Constrained principal components estimation of large approximate factor models

Rachida Ouysse*

October 31, 2019

Abstract

This paper proposes a constrained principal components (CnPC) estimator for efficient estimation of large-dimensional factor models when errors are cross-sectionally correlated and the number of cross-sections (N) may be larger than the number of observations (T). Although principal components (PC) method is consistent for any path of the panel dimensions, it is inefficient as the errors are treated to be homoskedastic and uncorrelated. The new CnPC exploits the assumption of bounded cross-sectional dependence, which defines Chamberlain and Rothschild's (1983) approximate factor structure, as an explicit constraint and solves a constrained PC problem. The CnPC method is computationally equivalent to the PC method applied to a regularized form of the data covariance matrix. Unlike maximum likelihood type methods, the CnPC method does not require inverting a large covariance matrix and thus is valid for panels with $N \geq T$. The paper derives a convergence rate and an asymptotic normality result for the CnPC estimators of the common factors. We provide feasible estimators and show in a simulation study that they are more accurate than the PC

*School of Economics, University of New South Wales, Sydney, 2052, Australia. Email: rouysse@unsw.edu.au.

estimator, especially for panels with N larger than T, and the generalized PC type estimators (Choi [2012]), especially for panels with N almost as large as T.

Keywords: High dimensionality, unknown factors, principal components, cross-sectional correlation, shrinkage regression, regularization, pseudo-out-of-sample forecasting

JEL Classification: C11, C13, C33, C53, C55

1 Introduction

Factor models constitute the dominant framework used across many disciplines for realistic parsimonious representations of the dynamic behavior of large panels of time series. Principal components estimators (PCEs) of the common factors can be easily computed in panels where the cross-sectional dimension N is large, and is possibly larger than the sample size T. PCEs are feasible for any path of the panel dimensions and are consistent for both N and T going to infinity [Forni et al., 2009, 2005, 2004, Bai, 2003, Bai and Ng, 2003, Stock and Watson, 2002a,b]. However, PCEs are not efficient in the presence of heteroskedasticity and/or dependence in the error term. Methods based on maximum likelihood (ML) and generalized least squares (GLS) type principal components depend on estimating a high-dimensional covariance matrix, which is a challenging problem in large systems ($N > T$) when errors are dependent and heteroskedastic. Generally, to impliment ML and GLS type methods, a feasible estimator of the covariance matrix is needed. A natural candidate is the sample covariance matrix. The sample covariance matrix behaves optimally if N is fixed and converges to the population covariance at a rate $T^{-1/2}$. However, when $N \to \infty$, the sample covariance matrix can behave very badly and for $N > T$ cannot be inverted. One common solution in the literature is to regularize the covariance matrix; see for example Fan et al. [2016] for an overview of this literature.

This article is related to a large literature on factor models and a much smaller literature on estimation when N is large and the errors are cross-sectionally dependent. Approxiamte factor models are consistently estimated using PC or ML methods. The fundamental result in the literature is that common factors can be consistently estimated for both N and T going to infinity, with no restrictions on the relative rates of convergence and under fairly general conditions on the time and cross-sectional dependence of the errors [Stock and Watson, 1998, 2002a,b, 2006, Bai and Ng, 2002, Bai, 2003, Kapetanios, 2010, Onatski, 2010]. These studies allow for an approximate factor structure but the error dependence dynamics do not enter the consistency result. Therefore, the idiosyncratic errors are treated as homoskedastic and as independent both in the cross-section and time dimensions. In general, although there exist well-established estimation procedures for static factor models, efficiency considerations have only received selective attention in the literature. Boivin and Ng [2006] documented, through an extensive simulation analysis, the potential effects of the presence of cross-sectional dependence on the small sample properties of the PCEs and their performance in forecasting. They find that "weighting the data by their properties when constructing the factors also lead to improved forecasts."Additionally they also find that with cross-correlated errors, the estimated factors may be less useful for forecasting when more series are available. The ML estimation provides a natural framework to account for heteroskedasticity and temporal dependence [Forni et al., 2004, 2009]. Doz et al. [2012] establish the properties of ML estimators for factor models in large panels of time series under heteroskedasticity. Breitung and Tenhofen [2011] propose a two-step GLS estimation that generalizes PC method to account for heteroskedasticity and serial correlation in a dynamic factor model with possibly large N. Choi [2012] considers efficient estimation using generalized least squares type PCEs to account for heteroskedasticity and dependence, but the framework requires $N < T$ to invert the sample covariance matrix. Bai and Li [2012] consider estimation of panels with

$N > T$ by maximizing the Gaussian-quasi likelihood method, but they consider the strict factor model with heteroskedastic but uncorrelated errors.

The literature is even more sparse in regards to efficiency considerations in large panels with large N (possibly larger than T) and cross-correlated errors. To our knowledge, Bai and Liao [2016] is the most relevant study for this article. Bai and Liao [2016] propose ML estimation with penalization of a large sparse covariance matrix. Their method is a joint estimation of the factors, their loadings and the covariance matrix, and is shown to be more efficient than PC or GLS type PC methods. Bai and Liao's (2016) paper is related to a growing literature on estimating large covariance matrices [Ledoit and Wolf, 2004, Bickel and Levina, 2008a,b, Ledoit and Wolf, 2012, Lam and Fan, 2009a]. Fan et al. [2011, 2013b] also used adaptive thresholding technique, as in Cai and Liu [2011], to estimate a sparse error covariance matrix in an approximate factor model using PC method when $N > T$. Advances in matrix theory have opened a new line of research into the consistent estimation of large matrices. Once a consistent estimate of the covariance matrix is achieved, a GLS type estimation or ML can be implemented in a two-step plug-in estimation approach. Regularisation and shrinkage methods amount to shrinking the off-diagonal elements of the error covariance matrix to zero. The key assumption these studies make is that the model is *sparse* or *conditionally sparse* (as in a factor model). The sparsity assumption of the error covariance matrix requires many off-diagonal elements to be zero or nearly zero. This assumption is slightly stronger than the original assumptions on cross-sectioanl dependence in Chamberlain and Rothschild [1983].

This article studies the efficient estimation of large-dimensional factor models, where N is large and is possibly larger than T, and where the errors are cross-sectionally dependent, while assuming the factors and the errors are stationary, and the error covariance matrix is time and cross-section separable. We propose a new PC-based estimation method, we denote as CnPC, that preserves the simple tractable implimen-

tation of PC method, while using infomation about the cross-sectional dependence in the data. Our estimation method solves a least squares problem, similar to the PC method, under a constraint derived from the assumption of bounded cross-sectional dependence in the sense of Chamberlain and Rothschild [1983]. The new estimators, we call constrained principal components estimators (CnPCEs), are computationally easily obtained by performing an eigenvalue decomposition to a regularized form of the data covariance matrix. The constrained estimation has a dual problem that can be cast as a shrinkage estimation, where the regularization is applied to the cross-sectional correlations in the data.

Our method differs from the various regularization techniques, proposed in the large covariance matrix literature, in that we move away from targeting the covariance matrix for estimation. As a result, we do not make the assumption of sparsity on its off-diagonal elements. Additionally, the constraint in the CnPC is less restrictive than sparsity, allowing for more flexible patterns in the cross-sectional dependence.

The CnPC method has an implicit effect of shrinking the cross-sectional correlations that is driven by the dynamics of the dependence in the data. The asymptotic properties of the CnPCEs of the common factors are derived using the existing techniques of Bai and Ng [2002], Bai [2003] and Choi [2012]. We derive a convergence rate for the CnPCEs to the population common factors and show an asymptotic normality result for appropriate path of (N, T). The CnPCEs are computationally more attractive (than ML-based estimators) because the estimation does not require (i) explicit assumption about the structure of sparsity of the covariance matrix, or (ii) estimation and inversion of large covariance matrices.

In finite samples, a simulation analysis shows that the CnPCEs have improved accuracy compared to PCEs and to GLS-PCEs. When applied to the problem of forecasting U.S. inflation and industrial production using the *diffusion indexes* framework of Stock and Watson [2002a], we find a relative improvement in accuracy of up to 6%

decrease in the out-of-sample mean square forecast errors for a ten-year rolling window forecasting exercise.

The rest of the paper is organized as follows. Section 2 reviews some results of the dynamic factor models and the method of PC. In Section 3, we introduce the new CnPCEs and provide feasible estimators. Section 4 establishes the asymptotic convergence result and the relative efficiency of CnPCEs. The small sample properties of the estimators are compared in Section 5 by means of Monte Carlo simulations. Finally, Section 5 concludes the article. Proofs are deferred to the Appendix.

Notation

The following notation is used throughout the paper: $E(.|Z_t)$ and $E_t(.)$ denote conditional expectation given variables in Z_t and given information at time t respectively. A' denotes the transpose of A, when $A = [a_{i,j}]$ is $q \times p$ matrix, $A' = [a_{j,i}]$ is of dimensions $p \times q$. $A \otimes B$ denotes the Kronecker product of matrices A and B, for $A = [a_{ij}]$ and $B = [b_{ij}]$, $A \otimes B = [a_{ij}B]$. A^{-1} denotes the inverse of a matrix A. ι_m is a m-vector of ones. I_m is an $m \times m$ identity matrix. $\text{diag}(A) = (a_{1,1}, a_{2,2}, ..., a_{n,n})$ when $A = [a_{i,j}]$. "vector"means column vector. For any positive number a, $[a]$ is the largest integer smaller than or equal to a. I(statement) is an indicator function that takes on value of 1 if 'statement' is true.

2 The Model and assumptions

Consider a sequence of random variables $\{X_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ which admits a static $r-$factor structure:

$$X_{it} = \lambda_i^{0\prime} F_t^0 + e_{it}, \tag{2.1}$$

where $F_t^0 = \{F_{kt}^0\}_{1 \leq k \leq r}$, is an $r-$vector of common factors at time t, $\lambda_i^0 = \{\lambda_{ik}^0\}_{1 \leq k \leq r}$ is the corresponding vector of factor loadings for cross-section unit i, and e_{it} is the

idiosyncratic term. Suppose that we observe X_{it} with $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Let us denote the data vector at time t, $(X_{1t}, \cdots, X_{Nt})'$, as $\underline{X}_t$, and denote its idiosyncratic component, $(e_{1t}, \cdots, e_{Nt})'$, as $\underline{e}_t$. The factor structure can be represented in vector form:

$$\underline{X}_t = \mathbf{\Lambda}^0 F_t^0 + \underline{e}_t, t = 1, \cdots, T, \tag{2.2}$$

where $\mathbf{\Lambda}^0$ denotes the $N \times r$ matrix of factor loadings with vector entries λ_i^0 for cross-section i, $\mathbf{\Lambda}^0 = \{\lambda_1^{0'}, \cdots, \lambda_N^{0'}\}$. In a more compact matrix notation,

$$\mathbf{X} = \mathbf{F}^0 \mathbf{\Lambda}^{\mathbf{0}'} + \mathbf{e},$$

where we denote the $T \times N$ data matrix as, $\mathbf{X} = [\underline{X}_1, \cdots, \underline{X}_T]'$, and denote the $T \times N$ matrix of idiosyncratic errors as, $\mathbf{e} = [\underline{e}_1, \cdots, \underline{e}_T]'$.

Let us denote the covariance matrix of the $N-$variate random variable $\underline{X}_t$ as Ψ_N, where $\Psi_N = \mathbf{E}(\underline{X}_t \underline{X}_t')$. The $r-$factor structure in (2.1) implies that Ψ_N admits the following decomposition:

$$\Psi_N = \mathbf{\Lambda}^0 \Sigma_F \mathbf{\Lambda}^{0'} + \Omega_N, \tag{2.3}$$

where $\Sigma_F = E\left(F_t^0 F_t^{0'}\right)$ is defined in Assumption A1, and the subscript N is explicit to indicate that the factor structure depends on the cross-section dimension. On the other hand, Ψ_N in (2.3) is time invariant, which is true under a stationarity assumption we make in this paper (see model assumptions below). Additionally, separability of the time and cross-sectional dependence, implies that the $NT \times NT$ covariance matrix of vec($\mathbf{e}$) is separable:

$$E\left[\text{vec}(\mathbf{e})\text{vec}(\mathbf{e})'\right] = \Omega_N \otimes \Theta_T,$$

where the $T \times T$ matrix Θ_T, captures the time dependence dynamics in the errors. In this paper, we assume that $\Theta_T = I_T$ and focus on the issue of capturing the dynamics of Ω_N in the estimation.

The approximate factor structure (Chamberlain and Rothschild [1983]) generalizes the strict factor model, which assumes a diagonal error covariance Ω_N, and allows for a more general covariance structure with both time and cross-sectional dependence in the errors. The existence and identification of the approximate factor structure requires some pervasiveness assumptions. Chamberlain and Rothschild [1983] show that if, as $N \to \infty$, the covariance matrix of the data has only r unbounded eigenvalues and the remaining eigenvalues remain constant, then there is a factor structure and it is unique [Brown, 1989, Connor and Korajczyk, 1993]. The dimension of the panel in Chamberlain and Rothschild's (1983) approximate structure can be large in both N and T. In fact, the high-dimensional aspect of N large is needed to achieve the desirable statistical properties of the estimators in an approximate factor model. One key assumption for identification of the approximate factor structure, is that the cross-sectional dependence of the errors is asymptotically weak. As the number of variables in the panel grows larger, the correlation between these variables becomes smaller. At the limit, when N goes to infinity, the correlation dies out, which ensures identification and consistent estimation of the column space of the factors [Stock and Watson, 2002a, Bai and Ng, 2002], as well as the inferential theory [Bai, 2003, Bai and Ng, 2003]. The consistency result is achieved even if the estimation method doesn't exploit features of the data, such as heterogeneity in the signal to noise ratio, and non-spherical error components.

The underlying assumptions of the approximate factor structure are standard in the literature. In particular the following assumptions are made, [Bai and Ng, 2002, Bai, 2003]

Assumption A1 (**Factors**). *$E\|F_t^0\|^4 < \infty$ and $\frac{1}{T}\sum_{t=1}^{T} F_t^0 F_t^{0\prime} \xrightarrow{p} \Sigma_F$ as $T \to \infty$, for an $r \times r$ non-random positive definite matrix Σ_F.*

Assumption A2 (**Factor Loadings**). *λ_i is either deterministic such that $\|\lambda_i^0\| < \overline{\lambda} < \infty$, or it is stochastic such that $E\|\lambda_i^0\|^4 < \overline{\lambda} < \infty$. In either case, $\|N^{-1}\sum_{i=1}^{T} \lambda_i^0 \lambda_i^{0\prime} -$*

$\Sigma_{\mathbf{\Lambda}}\| \xrightarrow{p} 0$ *as* $N \to \infty$ *for some* $r \times r$ *positive definite matrix* $\Sigma_{\mathbf{\Lambda}}$.

Assumption A3 (**Error term**). *There exists a positive constant* $M < \infty$*, such that for all* N *and* T*,*

1. $E(e_{it}) = 0, E|e_{it}|^8 \leq M$*;*
2. $E(\underline{e}'_s \underline{e}_t/N) = \gamma_N(s,t)$*,* $|\gamma_N(s,s)| \leq M$ *for all* s *and* $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_N(s,t)| \leq M$*;*
3. $E(e_{it}e_{jt}) = \tau_{ij,t}$ *with* $|\tau_{ij,t}| \leq |\tau_{ij}|$ *for some* τ_{ij} *and for all* t*; in addition,*

$$N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{ij}| \leq M;$$

4. $E(e_{it}e_{js}) = \tau_{ij,ts}$ *and* $(NT)^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{ij,ts}| \leq M$*;*
5. *For every* (t,s)*,* $E\left|N^{-1/2}\sum_{i=1}^{N}\left[e_{is}e_{it} - E(e_{is}e_{it})\right]\right|^4 \leq M$.

Assumption A4. *Weak Dependence between Factors and Idiosyncratic Errors:*

$$E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0 e_{it}\right\|\right) \leq M.$$

Assumption A2 allows the loading to be either deterministic or random. While the results in this paper are derived under deterministic λ_i, as noted in Bai and Ng [2002], the results can be extended to the case of stochastic λ_i, provided the loadings are independent of factors and idiosyncratic errors.

Assumptions A1 and A2 imply the existence of an r factor structure. The only observables are the data $\underline{X}_t, t = 1, \cdots, T$. The common factors $\mathbf{F}^0$, the loadings $\mathbf{\Lambda}^0$, and the covariance matrix Ω_N are all unknown population parameters. Additionally, the number of factors r is generally unknown. The literature has extensively studied the estimation of the number of common factors in panels of large dimensions. For

static approximate factor models like the one studied in this paper, Bai and Ng [2002] proposed a consistent estimator of r by minimizing information criteria that depend on both N and T. However, the finite sample properties of the estimated r may be sensitive to a prespecified threshold in the penalty function and to the maximum number in the support of r, especially in the presence of moderate to strong dependence of the errors. Alessi et al. [2010] revisits the penalty term in Bai and Ng's (2002) information criteria and adds a multiplicative tuning constant, based on Hallin and Liška's (2007) diverging eigenvalue method for generalized factor models. Examples of other estimators with improved finite sample properties are Onatski's (2010) "Edge Distribution" estimator and Ahn and Horenstein's (2013) "Eigenvalue Ratio"and "Growth Ratio"estimators. See also Onatski [2009] for inference about r, and Forni et al. [2000], Amengual and Watson [2007], and Bai and Ng [2007] for dynamic factor models. Recently, Bai and Ng [2017] propose a data-dependent penalty, based on shrinking singular values of the common components, that is more conservative in the presence of weak factors or measurement errors. Therefore, we take the number of factors r as known, as it can be consistently estimated using the aforementioned methods developed in the literature.

Under the regularity conditions in Assumptions **A1-A4** (Bai and Ng [2002], Stock and Watson [2002b]), the factors and factor loadings can be consistently estimated, as N and T are both large, by using the method of *asymptotic principal components* [Connor and Korajczyk, 1989]. Technically, PCE minimizes the objective function

$$V(\mathbf{\Lambda}, \mathbf{F}) \;=\; \mathrm{tr}\left[(\mathbf{X} - \mathbf{F}\mathbf{\Lambda}')'(\mathbf{X} - \mathbf{F}\mathbf{\Lambda}')\right], \tag{2.4}$$

by choosing the normalization $\mathbf{F}'\mathbf{F}/T = I_r$ and $\mathbf{\Lambda}^{0'}\mathbf{\Lambda}^0$ is a diagonal matrix with distinct entries [Bai and Ng, 2013]. The estimator has a simple interpretation in terms of the singular value decomposition of the sample covariance of the data. Consider the

spectral decomposition of the sample covariance matrix of $\mathbf{X}$, $\Psi_N = \frac{1}{T}\mathbf{X}'\mathbf{X}$:

$$\Psi_N \Gamma = \Gamma \Delta,$$

where $\Delta = \text{diag}(d_1, \cdots, d_N)$ is a diagonal matrix with d_l corresponding to the l^{th} highest eigenvalue of Ψ_N, and $\Gamma = (\varphi_1, \cdots, \varphi_N)$ is the matrix whose columns correspond to the normalized eigenvectors of Ψ_N. The normalized PCEs of $\mathbf{F}^0$ are $\widehat{F}_{k,t} = \frac{1}{\sqrt{d_k}}\varphi_k' \underline{X}_t$, for $k = 1, \cdots, r$ [De Mol et al., 2008]. Then $\hat{\mathbf{\Lambda}} = \mathbf{X}'\hat{\mathbf{F}}/T = \Gamma_{1:r}$ is a $N \times r$ matrix of estimated factor loadings, obtained as a least squares projection of $\mathbf{X}$ on $\hat{\mathbf{F}}$. One key result is that, while the restrictions identify the space spanned by the columns of $\mathbf{F}^0$ and the space spanned by the columns of $\mathbf{\Lambda}^0$, they do not necessarily identify the individual columns of $\mathbf{F}^0$ and $\mathbf{\Lambda}^0$.

The statistical properties of $\hat{\mathbf{F}}$ and $\hat{\mathbf{\Lambda}}$ are well studied in the literature. See for example Bai and Ng [2002] for consistency, and Bai [2003] for inferential theory. Referring to Bai and Ng's (2002) convergence result, the estimated factors $\hat{F}_t$ span the space of the true factors F_t^0 up to an orthogonal rotation H^k, meaning that

$$C_{NT}^2 \left(\frac{1}{T} \sum_{t=1}^{T} \| \hat{F}_t^k - H^{k'} F_t^0 \|^2 \right) = O_p(1),$$

where $H^k = \left(\mathbf{\Lambda}^{0'} \mathbf{\Lambda}^0 / N \right) \left(\mathbf{F}^{0'} \hat{\mathbf{F}} / T \right) \Delta_{1:r,1:r}$ and $C_{NT} = min\{\sqrt{N}, \sqrt{T}\}$.

Assumption A3.3 does not explicitly play a role, and is not operational, in the estimation of the factors and the loadings. The PCEs in an approximate factor model are the same as the PCEs in a strict factor model, where the error covariance matrix is diagonal and homoskedastic.

If Ω_N is known, a generalized least squares type principal components estimator

can be constructed by minimizing the objective function,

$$V_{\Omega_N}(\mathbf{\Lambda}, \mathbf{F}) \;=\; \text{tr}\left[\Omega_N^{-1}(\mathbf{X} - \mathbf{F}\mathbf{\Lambda}')'(\mathbf{X} - \mathbf{F}\mathbf{\Lambda}')\right]. \tag{2.5}$$

A similar GLS-PC estimator is studied by Choi [2012], although inspired by an ML approach, for the case of heteroskedastic errors, where $\Omega_N = \text{diag}[E(e_{1t}^2), \cdots, E(e_{Nt}^2)]$, and block diagonal cross-sectional dependence, where $\Omega_N = \Omega_1 \bigoplus \Omega_2 \cdots \bigoplus \Omega_n$ has n blocks. Breitung and Tenhofen [2011] consider a similar type estimation for dynamic factor models with heteroskedasticity and serial correlation. In the case of Ω_N unknown, the sample covariance matrix is used to propose feasible estimators. However, the sample covariance matrix is singular for high-dimensional systems with $N > T$, making these GLS-type estimators unfeasible.

Boivin and Ng [2006] propose a weighted PCE which minimizes the objective function $\sum_{i=1}^{N} w_{iT} \sum_{t=1}^{T}(X_{it} - \lambda_i' F_t)^2$, where choices of the weights include the following: (i) w_{iT} is the inverse of the diagonal element of $\widehat{\Omega}_{N,T}$, the sample covariance matrix estimated using data up to time T and, (ii) w_{iT} is the inverse of $N^{-1}\sum_{i=1}^{N} |\widehat{\Omega}_{N,T}(i,j)|$.

The random matrix literature has a rich body of work on estimating large dimensional covariance matrices [Fan et al., 2013a, 2016]. This literature makes a key assumption that the covariance matrix is sparse or conditionaly sparse, while using thresholding and penalized maximum likelihood for estimation [Cai and Liu, 2011]. Bai and Liao [2016] are perhaps the first to address efficiency in a high dimensional approximate factor model in the presence of heteroskedasticity and cross-sectional dependence in the errors. They apply Fan et al.'s (2013a) principal orthogonal component thresholding estimator to propose a two-step estimator, and Lam and Fan's (2009b) penalized likelihood to propose a joint estimation of $\mathbf{\Lambda}^0$ and a conditionally sparse Ω_N. In a recent contribution to this literature, Bai and Ng [2017] propose a minimum rank estimator by using ridge regressions and shrinking the singular values of the common

component.

3 The CnPC Estimator

Let us consider Assumption A3.3 of bounded cross-sectional correlation in an approximate factor structure. The eigenvalues of the error covariance matrix Ω_N in Chamberlain and Rothschild's (1983) factor model must be bounded. Under the assumption of (covariance) stationarity, $E(e_{it}e_{jt}) = \tau_{ij}$, all the eigenvalues of Ω_N are bounded by $\max_i \sum_{i=1}^{N} |\tau_{ij}|$. Thus Assumption A3.3 is implied by the assumption of $\sum_{i=1}^{N} |\tau_{ij}| \leq M$ for all i and all N as found in Bai and Ng [2002].

Let $\text{sgn}(a)$ denote the spatial sign function with $\text{sgn}(a) = |a|/a$ for $a \neq 0$ and $\text{sgn}(0) = 1$. Under the assumption of stationarity, Assumption A3.3 can be written as:

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \text{sgn}(\tau_{ij})\tau_{ij} \leq M_N, \tag{3.1}$$

where $\tau_{ij} = E(e_{it}e_{jt})$ and $e_{st} = X_{st} - \lambda_s' F_t$, for $s = i, j$, and where the bounding constant M is now indexed by N for the remainder of the paper.

The paper proposes a CnPC method that solves a PC problem under the constraint in (3.1):

$$\underset{\lambda_i, F_t}{\text{minimize}} \; (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} e_{it}^2 \tag{3.2}$$

$$\text{s.t} \; \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \text{sgn}(\tau_{ij})\tau_{ij} \leq M_N \tag{3.3}$$

Let $\mathcal{S}$ be $N \times N$ matrix with elements $[\mathcal{S}_{ij}]$ defined as,

$$\begin{aligned} \mathcal{S}_{i,i} &= 0 \\ \mathcal{S}_{i,j} &= \text{sgn}\left(\tau_{ij}\right) \text{ for } i \neq j. \end{aligned}$$

3.1 The CnPC estimator with known $\mathcal{S}$

In this next section, we proceed with the working assumption that $\mathcal{S}$ is known to derive the CnPC estimator. This is an unfeasible estimator since in practice, the sign of correlation between any two variables is generally unknown. However, it is plausible that in many applications, institutional knowledge and theory may provide information about the direction of co-variation between variables without knowledge about the strength of the relationship. We show in Lemma2, below, that a consistent estimator of $\mathcal{S}$ can be obtained.

Let $\mathcal{L}_1(\mathbf{F},\Lambda) = \frac{1}{T}\sum_{t=1}^{T} \underline{e}_t'\underline{e}_t$ and $\mathcal{L}_2(\mathbf{F},\Lambda) = \frac{1}{NT}\sum_{t=1}^{T} \underline{e}_t'\mathcal{S}\underline{e}_t - M_N$. The optimization in (3.2)-(3.3) can be written as:

$$\underset{\mathbf{\Lambda},\mathbf{F}}{\text{minimize}}\ \{\mathcal{L}_1(\mathbf{\Lambda},\mathbf{F})|\mathcal{L}_2(\mathbf{F},\mathbf{\Lambda}) \leq 0\}, \tag{3.4}$$

under the normalization restrictions that $T^{-1}\sum_{t=1}^{T} F_t F_t' = I_r$ and $N^{-1}\sum_{i=1}^{N}\lambda_i\lambda_i'$ is an $r \times r$ diagonal matrix with distinct entries. This optimization problem can be solved using standard Kuhn-Tucker's theorem. Treating the system in (3.4) as a convex programming problem, the Lagrangian is

$$\mathcal{L}(\mathbf{\Lambda},\mathbf{F},\mu) = \frac{1}{N}\mathcal{L}_1(\mathbf{\Lambda},\mathbf{F}) + \mu_{NT}\mathcal{L}_2(\mathbf{F},\mathbf{\Lambda}). \tag{3.5}$$

The matrix $\mathcal{S}$ has diagonal elements equal to zero and off-diagonal elements that are equal to either 1 or -1. The Lagrangian is similar to that of a shrinkage regression where the cross-sectional correlations are shrunk towards zero. The tuning parameter μ_{NT} represents the cost/penalty for deviation of the solution from (3.1) and thus plays the role of a shrinkage factor.

Proposition 1. *The constrained principal components estimator for* $\mathbf{F}^0$*, denoted* $\hat{\mathbf{F}}$*, which solves (3.5), is* $\sqrt{T}$ *times the matrix consisting of the eigenvectors corresponding*

to the r largest eigenvalues of the matrix $\frac{\mathbf{X}\mathcal{A}_N\mathbf{X}'}{NT}$, where $\mathcal{A}_N = I_N + \mu_{NT}\mathcal{S}$, and μ_{NT} is the Lagrange multiplier parameter. The CnPC estimator for $\mathbf{\Lambda}^0$, denoted $\hat{\mathbf{\Lambda}}$ is given by $\hat{\mathbf{\Lambda}} = \frac{1}{T}\mathbf{X}\hat{\mathbf{F}}$.

Assumption A5. *There exists a positive constant $D < \infty$, such that for all N and T,*

1. *let $E(\underline{e}_s'\mathcal{S}\underline{e}_t/N) = \varrho_N(s,t)$, then $\sum_{s=1}^{T}|\varrho_N(s,t)| \leq D$ for all t.*
2. *for every t, s, and N, assume that $E\left|N^{-1/2}\left[\underline{e}_s'\mathcal{S}\underline{e}_t - E\left(\underline{e}_s'\mathcal{S}\underline{e}_t\right)\right]\right|^4 \leq D$;*
3. *for any t and N, there exists a positive constant $M < \infty$ such that $E\left\|\frac{1}{\sqrt{N}}\mathbf{\Lambda}^{0\prime}\mathcal{S}\underline{e}_t\right\|^2 \leq D$.*

Theorem 3.1. *For any fixed (known) $r \geq 1$, there exists a suitable $(r \times r)$ full rank rotation matrix $\mathcal{H}$ such that under Assumption A1-A5*

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - \mathcal{H}'F_t^0\right\|^2 = O_p(\delta_{NT}^{-2}) + O_p(\mu_{NT}^{-2}\delta_{NT}^{-2}),$$

where $\mathcal{H} = \left(\frac{\mathbf{\Lambda}'\mathcal{A}_N\mathbf{\Lambda}}{N}\right)\left(\frac{\mathbf{F}'\hat{\mathbf{F}}}{T}\right)V_{NT}^{-1}$. Or equivalently,

$$\omega_{NT}^2\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - \mathcal{H}'F_t^0\right\|^2\right) = O_p(1),$$

where $\delta_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$ and $\omega_{NT} = \min\{\delta_{NT}, \delta_{NT}\mu_{NT}\}$.

In Theorem 3.1, the time average of squared deviations between the CnPC estimator and those that lie in the true factor space goes to zero as $N, T \to \infty$. The rate of convergence depends on the panel structure but also on the regularization factor μ_{NT}. When $\mu_{NT} = O(1)$, equivalent to $h = 0$ in Proposition2 below, the CnPC estimator of $\hat{F}$ is the principal components estimator of the factor space consisting of the eigenvectors corresponding to the r largest eigenvalues of $\mathbf{X}\mathbf{X}'/T$ [Stock and Watson, 2002a, Bai and Ng, 2002, Bai, 2003]. In this case, Theorem 3.1 implies the same rate of convergence

as in Bai and Ng [2002] which is equal to δ_{NT} and is determined by the smaller of N or T.

Theorem 3.1 establishes conditions under which the convergence of the CnPC estimator is faster/slower than that of the ordinary PCEs.

Proposition 2. *Let $\mu_{NT} = \delta_{NT}^{-h}$, then the rate of convergence in Theorem 3.1 is:*

(i) $\omega_{NT}^2 = \delta_{NT}^{2(1-h)}$ *for* $h > 0$,

(ii) $\omega_{NT}^2 = \delta_{NT}^2$ *for* $h \leq 0$.

In the case of $h > 0$, $\omega_{NT}^2 < \delta_{NT}^2$ and thus the CnPC estimator converges (in the sense of Theorem 3.1) to factors that lie in the true factors space at a rate slower than Bai and Ng's (2002) ordinary CPEs. The two methods imply a different rotation matrix $\mathcal{H}$, which means the convergence is towards a different rotation of the space spanned by the true factors. Thus the estimated factor spaces are not directly comparable.

Lemma 1. *Assume in addition that* $max_{1\leq t\leq T} \sum_{s=1}^{T} \gamma_N(s,t)^2 \leq M_N$ *for some* $M < \infty$ *uniformly in* t, *then*

$$\omega_{NT}^2 \left\| \hat{F}_t - \mathcal{H}' F_t^0 \right\|^2 = O_p(1).$$

The proof is similar to that of Theorem 3.1.

3.2 A consistent estimator for $\mathcal{S}$

The population pair-wise signs of the cross-sectional correlations, $\mathcal{S}_{ij}, i \neq j$ are generally unknown. The information we require is about the direction of association between the two cross-sectional units i and j. This does not necessarily require estimating the full covariance/correlation matrix. A sufficient statistic that measures the (conditional) ordinal association between $\mathbf{X}_i$ and $\mathbf{X}_j$ can be used to construct an estimator for $\mathcal{S}_{ij}$.

An estimate of $\mathcal{S}_{ij}$ can be defined using the estimated parameters of the model,

$$\begin{aligned}\hat{\mathcal{S}}_{ij} &= \mathcal{S}\left(\hat{\lambda}_i, \hat{\lambda}_j, \hat{F}_t\right) \\ &= \operatorname{sgn}\left[\widehat{E}\left(\hat{e}_{it}\hat{e}_{jt}\right)\right] \\ &= \operatorname{sgn}\left[\frac{1}{T}\sum_{t=1}^{T}\left(X_{it} - \hat{\lambda}_i'\hat{F}_t\right)\left(X_{jt} - \hat{\lambda}_j'\hat{F}_t\right)\right]\end{aligned}$$

Instead of estimating a covariance/correlation matrix, a pair-wise estimator for τ_{ij} can be computed only from the i^{th} and j^{th} cross sections. This is a faster and better strategy in high-dimensions with possibly sparse systems ($N > T$), as per Dürre et al. [2015]. Consider the sample moment estimator, $\hat{\tau}_{ij}$, for the population τ_{ij}:

$$\hat{\tau}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}\hat{e}_{jt},$$

where $\hat{e}_{kt} = X_{kt} - \hat{C}_{kt}$, where the common components estimator, $\hat{C}_{kt} = \hat{\lambda}_k'\hat{F}_t$, for $k = 1, \cdots, N$. In order to make Assumption (**A3.3**) operational, the population moments τ_{ij} are replaced by the sample moments $\hat{\tau}_{ij}$, and $\operatorname{sgn}(\tau_{ij})$ by $\operatorname{sgn}(\hat{\tau}_{ij})$.

Lemma 2 (Consistency of τ_{ij} and $\operatorname{sgn}(\hat{\tau}_{ij})$)**.** *Under assumptions A1-A4, as $T, N \to \infty$ we have*

i. $\hat{\tau}_{ij}$ converges to τ_{ij} at a rate $O_p\left(\frac{1}{T^{1/4}}\right) + O_p\left(\frac{1}{\delta_{NT}}\right)$

ii. For $\hat{\tau}_{ij} \neq 0$, plim $sgn(\hat{\tau}_{ij}) = sgn(\tau_{ij})$

3.3 Regularisation and thresholding

The largest eigenvalue of Ω_N is bounded by $\max_i \sum_{i=1}^{N} |\tau_{ij}|$, where $\tau_{ij} = E(e_{it}e_{jt})$, (Boivin and Ng [2006]). Under Assumption A3.3, there should exist a τ_N such that $\sum_{j=1}^{N} |\tau_{ij}| \leq \tau_N < \infty$ for all i and N. This assumption is vital in the development of the approximate factor structure theory. However, there is no indication as to how much

cross-correlation is permitted in practice. Boivin and Ng [2006] use $\hat{\tau}^* = \max_i \hat{\tau}_i^*/N$, where $\hat{\tau}_i^* = \sum_{j=1}^{N} |T^{-1} \sum_{t=1}^{T} \hat{e}_{it}\hat{e}_{jt}|$ as indicator for τ_N/N, which should be small and decrease with N.

The use of thresholding parameters to regularise the cross-sectional dependence is an implication of the requirement of sparsity/conditional sparsity of Ω_N, which is one key assumption in the literature for estimating high-dimensional covariance matrices. See Bickel and Levina [2008a], Cai and Liu [2011], Fan et al. [2016]. Sparsity requires that many of the cross-correlations are zero which translates to a sparse condition that requires the quantity:

$$m_T = \max_{i \leq N} \sum_{j \leq N} I(\tau_{ij} \neq 0), \tag{3.6}$$

to be bounded or grow slowly as $N \longrightarrow \infty$. That is $m_T = o(f(N))$, where $f(N) = o(N)$. This condition is a special case of a more general sparsity assumption that requires $m_T = \max_{i \leq N} \sum_{j=1}^{N} |\tau_{ij}|^q$, for some $0 \leq q < 1$, to be bounded. Condition (3.6) is stronger than Assumption A3.3 because it requires that the idiosyncratic components are mostly uncorrelated.

In specific terms, choices of M_N in A3.3 can be determined depending on the thresholding mechanisms. Given that $|\tau_{ij}| \leq 1, i, j \leq N$, it follows that

$$\max_i \sum_j |\tau_{ij}| \leq \max_i \sum_j I(\tau_{ij} \neq 0),$$

and $\tau^* \leq m_T$. Following Fan et al. [2011], it is sufficient to assume that for a fixed r

$$M_N \;\; = \;\; o\left(\left[\frac{T}{\log N} \right]^{0.5} \right), \tag{3.7}$$

where as in their notation $M_N \equiv \sum_{i \leq N} m_T/N$.

The CnPC estimation, the estimated factors and loadings are a function of the

threshold parameter M_N as it controls the 'shrinkage'parameter μ_{NT}, which is estimated from the concentrated objective function $\mathcal{L}(\mu_{NT})$. There is a correspondence between μ_{NT} and M_N: μ_{NT} increases as M_N decreases for given N, T. This correspondence is, however, not a function. Although positive values of μ_{NT} correspond to a single value of M_N, the value $\mu_{NT} = 0$ relates to all M_N in $\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|E(\hat{e}(0)_{it}\hat{e}(0)_{jt})|, \infty\right)$, where $\hat{e}(0)_{it}$ and $\hat{e}(0)_{jt}$ are the residuals obtained in the PC estimation method. If the factor structure is strict, there is no need for shrinkage.

Let us denote as M_0 the amount of cross-sectional dependence in the sample covariance matrix, estimated using the PC residuals,

$$M_0 = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|T^{-1}\sum_{t=1}^{T}\hat{e}(0)_{it}\hat{e}(0)_{jt}\right|.$$

It is then the case that values of M_N that are smaller than M_0 will have an effect of an increase in shrinking the off-diagonal elements $\hat{\tau}_{ij}$.

The complementary slackness conditions are used to compute an estimate $\hat{\mu}_{NT}$ for the Lagrange multiplier. If the constraints are not binding and $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|E(e_{it}e_{jt})| \leq M_N$, then the constrained maxima are the ordinary PC solution $\left(\hat{F}, \hat{\mathbf{\Lambda}}, 0\right)$. On the other hand, if $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|E(e_{it}e_{jt})| > M_N$, then by the complementary slackness we must have $\hat{\mu}_{NT} > 0$ and $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|E(e_{it}e_{jt})| = M_N$.

If M_N is greater than or equal to the $L_{1,1}-$norm of the PC sample covariance matrix, M_0, then the PCE estimator is unchanged by the proposed regularization. For smaller values of M_N, the CnPC has an effect of shrinking the cross-sectional correlations towards the origin, in the $L_{1,1}$ sense.

In this paper, we consider a grid of values for M_N with a support from zero to M_0, which is computed using the residuals of a first stage PCE. The grid is indexed by a parameter, we call m, representing the percentage shrinkage applied relative to the unconstrained PCE. For example, a value of $m = 0.1$ means that the CnPC applies a

threshold that is equal to 10% the value of M_0, implying a very high level of shrinkage and a tightening of the constraint. One way to make a single choice of m is cross-validation. A 'best' value of m corresponds to the grid value that minimizes a specified risk/objective function. The risk can be measured in terms of fit of the estimated factors $\hat{F}$ and/or in terms of prediction error for factor based forecasts. In the simulation analysis, we present the path of solutions over the grid, as well as show some results for a selected best m.

It is worth noting that there is a dual problem to the CnPC problem that we solve in this paper. A closely related optimization problem to CnPC regression in (3.5) is the constrained regression

$$\underset{\lambda_i, F_t}{\text{minimize}} \; (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} e_{it}^2 + \kappa_{nt} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |E(e_{it} e_{jt})| \tag{3.8}$$

Problems (3.5) and (3.8) are equivalent (Osborne et al. [2000]). For a given $\kappa_{nt}, 0 \leq \kappa_{nt} < \infty$, there exists a $M_N \geq 0$ such that the two problems share the same solution, and vice versa. In (3.8), the parameter κ_{nt} is easily interpreted as a shrinkage/regularization parameter applied to a large cross-section of correlation parameters. The Lagrange multiplier μ_{NT} is the price of deviation from the bounded cross correlation constraint imposed by the approximate factor structure. The two parameters are exchangeable for all practical purposes.

4 Limiting distributions of CnPC estimators

In this section, we study the asymptotic distributions of the proposed CnPCEs. These estimators are compared to the ordinary PCEs [Bai, 2003] and the generalized PCEs [Choi, 2012].

Assumption A6. *Moments and Central Limit Theorem*

1. *for any t, N and T, there exists an $D < \infty$ such that*

$$E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T} F_s^0\left[\underline{e}_s'\mathcal{S}\underline{e}_t - E\left(\underline{e}_s'\mathcal{S}\underline{e}_t\right)\right]\right\|^2 \leq D;$$

2. *for any N and T, there exists an $D < \infty$ such that*

$$E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T} \mathbf{\Lambda}^{0'}\mathcal{S}\underline{e}_s\mathbf{F}_s^{0'}\right\|^2 \leq D;$$

3. *for each t, as $N \to \infty$,*

$$\frac{1}{\sqrt{N}}\mathbf{\Lambda}^{0'}\underline{e}_t \xrightarrow{d} N(0, \Psi_t)$$

where $\Psi_t = \lim_{N\to\infty}\frac{1}{N}\mathbf{\Lambda}^{0'}E(\underline{e}_t\underline{e}_t')\mathbf{\Lambda}^0$;

4. *for each i, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{F}_t^0 e_{it} \xrightarrow{d} N(0, \Phi_i),$$

where $\Phi_i = plim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E(F_t^0F_t^{0'}e_{it}e_{is})$.

Assumption A7 (Factor Loadings*). $\|N^{-1}\mathbf{\Lambda}^{0'}\mathcal{A}_N\mathbf{\Lambda}^0 - \Sigma_{\mathbf{\Lambda}*}\| \to 0$ *as* $N \to \infty$ *for some* $r \times r$ *positive definite matrix* $\Sigma_{\mathbf{\Lambda}*}$.

Assumption A8. *The eigenvalues of the* $r \times r$ *matrix* $(\Sigma_{\mathbf{\Lambda}*} \cdot \Sigma_{\mathbf{F}})$ *are distinct.*

As in Bai [2003], Assumption A8 is needed to guarantee a unique limit for the the matrix $\hat{\mathbf{F}}'\mathbf{F}^0/T$ because the factors are only identified up to an orthogonal transformation $\mathcal{H}$.

Proposition 3. *Under Assumptions A1-A5 and Assumption A8, the matrix $\mathcal{Q}$ defined as:*

$$\mathcal{Q} = plim_{T,N\to\infty}\frac{\hat{\mathbf{F}}'\mathbf{F}^0}{T},$$

is invertible and is given by $\mathcal{Q} = V^{1/2}\Psi\Sigma_{\mathbf{\Lambda}^*}^{1/2}$*, where* V *is a diagonal matrix of largest* r *eigenvalues of* $\Sigma_{\mathbf{\Lambda}^*}\Sigma_{\mathbf{F}}\Sigma_{\mathbf{\Lambda}^*}$*, and* Ψ *is the corresponding eigenvector matrix.*

The proof is similar to that of Proposition 1 in Bai [2003].

Theorem 4.1. *Suppose that Assumptions A1-A8 hold. In addition, if* $\frac{\sqrt{N}}{T\mu_{NT}} \to 0$*, then the CnPC estimator* $\hat{F}_t$ *has a limiting distribution,*

$$\sqrt{N}\left(\hat{F}_t - \mathcal{H}'F_t^0\right) \xrightarrow{d} N(0, V^{-1}\mathcal{Q}\Psi_t\mathcal{Q}'V^{-1}). \tag{4.1}$$

The main motivation of this paper is to improve on the existing estimators in terms of efficiency. The ordinary PCEs $\hat{F}_{t,PC}$ have asymptotic distribution (Bai's (2003) Theorem 1):

$$\sqrt{N}\left(\hat{F}_{t,PC} - H'F_t^0\right) \xrightarrow{d} N(0, V_{PC}^{-1}Q_{PC}\Psi_t Q'_{PC}V_{PC}^{-1}), \tag{4.2}$$

where $Q_{PC} = \Sigma_{\mathbf{\Lambda}}^{-1/2}\Upsilon_{PC}V_{PC}^{1/2}$, Υ_{PC} is the matrix of eigenvector of $\Sigma_{\mathbf{\Lambda}}^{1/2}\Sigma_F\Sigma_{\mathbf{\Lambda}}^{1/2}$, and $V_{PC} = Q_{PC}\Sigma_{\mathbf{\Lambda}}Q'_{PC}$.

In terms of efficiency, it is not clear how the asymptotic covariance matrix in (4.1) compares to the asymptotic covariance matrix in (4.2). The CnPC and PCE estimate different objects since the rotation matrices H and $\mathcal{H}$ are generally not identical.

Let us consider the simple case of a one factor model with $r = 1$. This is an interesting case where H and $\mathcal{H}$ are identical and equal to the scalar $\Sigma_{\mathbf{F}}^{-1/2}$. In this case, the PCEs and the CnPCEs are estimating the same object $\mathbf{F}_t/\sqrt{\Sigma_{\mathbf{F}}}$.

Corollary 1. *Suppose that Assumptions A1-A8 hold and* $r = 1$*, then the CnPCEs are more efficient than the PCEs with a ratio of (asymptotic) variances, for* $t = 1, \cdots, T$*:*

$$\frac{V\left(\hat{F}_{t,PC}\right)}{V(\hat{F}_{t,CnPC})} = \left(1 + \mu_{NT}\ plim\ \frac{\mathbf{\Lambda}'\mathcal{S}\mathbf{\Lambda}}{N}\right)^2 \geq 1.$$

In the case of one-factor model (of $r = 1$), $\Upsilon = \Upsilon_{PC} = 1$, $\mathcal{Q} = Q_{PC} = \Sigma_F^{-1/2}$, and the CnPC estimator $F_{t,CnPC}$

$$\hat{F}_{t,CnPC} \backsimeq \frac{F_t^0}{\sqrt{\Sigma_{\mathbf{F}}}} + \frac{1}{\sqrt{N}} N\left(0, \frac{1}{\Sigma_{\mathbf{F}}} \Sigma_{\mathbf{\Lambda}*}^{-1} \Psi_t \Sigma_{\mathbf{\Lambda}*}^{-1}\right) \tag{4.3}$$

and the PCEs have

$$\hat{F}_{t,PC} \backsimeq \frac{F_t^0}{\sqrt{\Sigma_{\mathbf{F}}}} + \frac{1}{\sqrt{N}} N\left(0, \frac{1}{\Sigma_{\mathbf{F}}} \Sigma_{\mathbf{\Lambda}}^{-1} \Psi_t \Sigma_{\mathbf{\Lambda}}^{-1}\right), \tag{4.4}$$

where

$$\Sigma_{\mathbf{\Lambda}*} = \Sigma_{\mathbf{\Lambda}} + \mu_{NT} \text{ plim } \frac{\mathbf{\Lambda}' \mathcal{S} \mathbf{\Lambda}}{N} \geq \Sigma_{\mathbf{\Lambda}}, \tag{4.5}$$

because $\mathcal{S}$ is positive definite and $\mu_{NT} \geq 0$.

5 Monte Carlo Simulations

5.1 Simulations designs

This section presents the Monte Carlo experiments designed to study the small sample properties of the proposed CnPC estimator and their performance relative to the ordinary PCEs in the presence of cross-correlated errors. The experimental design for the Monte Carlo simulation adopts the same covariance structure as in Boivin and Ng [2006]. Let the total number of cross-sections N be divided into three groups of sizes N_1, N_2 and N_3 such as $N = N_1 + N_2 + N_3$. Let the errors u_{it} be the building blocks for the errors dynamics with $u_{it} \sim N(0, 1), i = 1, \cdots, N$, and construct the errors e_{it} where

- $e_{it} = \sigma_1 u_{it}$ for the first block of N_1 time series,

- $e_{it} = \sigma_2 u_{it}$ for the second block of N_2 time series,
- $e_{it} = \sigma_3 \tilde{e}_{it}, \tilde{e}_{it} = u_{it} + \sum_{j=1}^{C} \rho_{ij} u_{jt}$ for the third block of N_3 time series.

The errors within the first block of N_1 series are mutually uncorrelated and homoskedastic, the errors within the second block of N_2 series are also mutually uncorrelated and homoskedastic but their variance differs from the variance of the first block of time series, $\sigma_2^2 > \sigma_1^2$. Cross-correlation is introduced in the third block of N_3 series. The series in block 3 are correlated with a proportion C of block 1 of N_1 cross-sections. The lower first block of the covariance matrix, Ω_{13}, therefore has $C \cdot N_3$ non-zero elements. The correlation coefficients ρ_{ij} denote cross-correlation of series $i \in \{1, N_1\}$ and $j \in \{N_1 + N_2 + 1, N\}$ and is drawn from a uniform distribution $U[0.05, 0.7]$. The error variance in the third group is $\sigma_3^2 = \sigma_1^2$. The error covariance matrix takes the following form:

$$
\begin{aligned}
\Omega_{ii} &= \sigma_1^2, && 1 \leq i \leq N_1 \\
\Omega_{ii} &= \sigma_2^2, && N_1 + 1 \leq i \leq N_1 + N_2 \\
\Omega_{ii} &= \sigma_3^2, && N_1 + N_2 + 1 \leq i \leq N \\
\Omega_{ij} &= 0, && 1 \leq i, j \leq N_1 + N_2 \\
\Omega_{ij} &= \sigma_1 \sigma_3 \rho_{ij}, && i \leq C, N_1 + N_2 + 1 \leq j \leq N.
\end{aligned}
$$

The common factors and their loadings are fixed throughout the simulation, which corresponds to an analysis conditional on $\mathbf{F}^0$ and $\mathbf{\Lambda}^0$. The number of factors r is known and fixed. We consider two values of r in the data generating process: r equals 1 for a small factor structure and r equals 4, representing a common number found in the empirical literature. The panel dimension takes combinations of $T = 50, 100$, and $N = 50, 100, 150$. Data are generated through $X_{it} = \sum_{m=1}^{r} \lambda_{im}^0 F_{mt}^0 + e_{it}$. The Monte Carlo results are based on $L = 2,000$ replications.

For each replication $l = 1, \cdots, L$, the Monte Carlo experiment is carried out as follows.

(i) Compute the ordinary PCEs of $\hat{\mathbf{F}}^{(l)}_{PCE}$, $\hat{\mathbf{\Lambda}}^{(l)'}_{PCE}$ and the estimated errors $\hat{\mathbf{e}}^{l}_{PCE} = \mathbf{X}^{(l)} - \hat{\mathbf{\Lambda}}^{(l)'}_{PCE}\hat{\mathbf{F}}^{(l)}_{PCE}$. Using the sample correlation between cross-sections i and j, $\hat{\tau}^{l}_{ij,PCE} = \hat{\mathbf{e}}^{l'}_{i,PCE}\hat{\mathbf{e}}^{l}_{j,PCE}/T$, construct an estimate for the elements of the sign matrix, $\hat{\mathcal{S}}^{(l)}_{ij}$, for $i \neq j$ and $i, j = 1, \cdots, N$.

(ii) Given a value of $M = m \cdot M_0$, where $m \in [0, 1]$, compute $\left(\hat{\mathbf{F}}^{(l)}, \hat{\mu}^{l}_{NT}\right)$:

(a) Begin with a starting value $\mu_{NT} = \mu_0$, here we take $\mu_0 = 0.5\sqrt{tr(\hat{\mathbf{e}}'\hat{\mathbf{e}})/tr(\hat{\mathbf{e}}'\mathcal{A}_N\hat{\mathbf{e}})}$ using the first stage PCEs and $\mathcal{A}_\mu = I_N - \mu\hat{\mathcal{S}}$. Find the solution to the dual objective function $\mathcal{L}(\mu)$:

$$\hat{\mu}_{NT} = \arg\max_{\mu}(NT)^{-1}\left[\text{tr } \mathbf{X}\mathcal{A}_\mu\mathbf{X}' - \text{tr } \hat{\mathbf{F}}'_\mu\mathbf{X}\mathcal{A}_\mu\mathbf{X}'\hat{\mathbf{F}}_\mu\right] - M, \qquad (5.1)$$

where $\hat{\mathbf{F}}_\mu$ is $\sqrt{T}$ times eigenvectors corresponding to the largest r eigenvalues of $\Psi_{N,\mu} = \frac{1}{T}\mathbf{X}'\mathcal{A}_\mu\mathbf{X}$. This is iterated to convergence and to optimal values $\hat{\mathbf{F}}^{(l)}, \hat{\mu}^{(l)}_{NT}$.

(b) Compute the CnPC estimator for the loadings as a linear projection of $\mathbf{X}$ on $\hat{\mathbf{F}}^{(l)}$: $\hat{\mathbf{\Lambda}}^{(l)} = \frac{1}{T}\mathbf{X}'\hat{\mathbf{F}}^{(l)}$.

(iii) Compute the following measures of performance.

- *Percentage explained variation.* Boivin and Ng [2006] use the percentage of variation in the true factors captured by the estimated structure,

$$S^{(l)}_{\hat{\mathbf{F}},\mathbf{F}^0} = \frac{\text{tr}\left(\mathbf{F}^{0\prime}\hat{\mathbf{F}}^{(l)}\left(\hat{\mathbf{F}}^{(l)\prime}\hat{\mathbf{F}}^{(l)}\right)^{-1}\hat{\mathbf{F}}^{(l)\prime}\mathbf{F}^0\right)}{\text{tr}(\mathbf{F}^{0\prime}\mathbf{F}^0)}.$$

- *Small sample bias.* The estimated factors and the true factors are not directly comparable. The estimated factors span a transformation of the true

factors. In comparing the small sample bias of the CnPC estimator and the benchmark PCEs, one has to account for the differences in the rotation matrices H and $\mathcal{H}$. We compute the small sample bias of the (rotated) factors $\tilde{F}_t \equiv \mathcal{H}^{-1}\hat{F}_t$:

$$\text{bias}^{(l)} = \frac{1}{L}\sum_{l=1}^{L}\tilde{F}_{tk}^{(l)} - F_{tk}^{0}, \tag{5.2}$$

for $k = 1$ and $t = 1, [T/2], T$.

- *Empirical mean squared errors (MSEs).* For each $\hat{F}_t^{(l)}$, we compute

$$MSEs^{(l)} = r^{-1}\left\|\hat{F}_t^{(l)} - F_t^{0(l)}\right\|^2. \tag{5.3}$$

5.2 The Diffusion Index framework

Consider the forecasting model whereby we are interested in the h-steps ahead forecast of a series y_t. In the presence of many predictors, the Diffusion Index of Stock and Watson [2002a] proposes using a small number of common factors, also called indexes, that are extracted from the matrix of predictors $\mathbf{X}$, to forecat a series y. The framework offers a solution to dimensionality and turns it from curse to blessing, [Stock and Watson, 2002b, Stock, 2005]. The data generating process for the forecast target series is:

$$y_{t+h} = \beta_0 + \sum_{j=1}^{r}\beta_j\mathbf{F}_{jt}^{0} + \epsilon_{t+h} \equiv y_{\mathbf{F}^0,t+h|t} + \epsilon_{t+h},$$

where $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, and σ_ϵ^2 is chosen such that the R^2 of the forecasting equation is κ_y. The unfeasible diffusion index forecast is $\hat{y}_{\mathbf{F}^0,t+h|t}$, which only requires estimation of β. The feasible diffusion index forecast is denoted as $\hat{y}_{\hat{\mathbf{F}},t+h|t}$, which requires estimation of both the factors and β. A forecast using the observed N series is not feasible if N is large. However, one can use the factor structure of X_{it} in equation (2.1) and use

$F_t^0 \equiv \{F_{jt}^0\}_{j=1}^r$ to account for the important drivers of the common variation in $\mathbf{X}$:

$$\hat{y}_{F^0,t+h|\mathcal{I}_t} = \hat{\beta}_0 + F_t^{0\prime}\hat{\beta}. \tag{5.4}$$

This forecast is unfeasible since the true factors F_t^0 are unobserved. Given estimates $\hat{F}_{t,N} \equiv \{\hat{F}_{jt,N}\}_{j=1}^{\hat{r}}$, using the data from the N series and conditional on information at time $\mathcal{I}_t$, a feasible factor augmented forecast is constructed as

$$\hat{y}_{\hat{F}_{t,N},t+1|\mathcal{I}_t} = \hat{\beta}_0 + \hat{F}_{t,N}'\hat{\beta}. \tag{5.5}$$

The feasible forecast depends on the properties of both the estimated β's and the 'generated' regressors $\hat{F}_{t,N}$.

The performance of forecasts is evaluted using a pseudo-out-of-sample forecasting exercise with a ten years rolling window. The simulated series are divided into in-sample data and out-of-sample data. The first T observations of the series make up the in-sample data. The starting point for the out-of-sample exercise is $t = T$. At each time $t = T, \cdots, T + J - h$, estimation is carried out using the last 10 years of data, and point forecasts for the target series at time $T + h$ are formed. Next, the sample is rolled forward by one observation and the same exercise is repeated for J periods. At the end a series of J pseudo-out-of-sample forecasts is available, with a choice of $J = 120$. The panel dimensions in this experiment are $T = 120$ and $N = 131$ to reflect those commonly used in macroeconomic forecasting. We compute the empirical mean-squared-forecast errors (MSFE) as in Boivin and Ng [2006]

$$MSFE_{\hat{y}^{\hat{\mathbf{F}}},\hat{y}^{\mathbf{F}^0}} = \frac{1}{J}\sum_{t=T}^{T+J-1}\left(\hat{y}_{\mathbf{F}^0,t+1|t} - \hat{y}_{\hat{\mathbf{F}},t+1|t}\right)^2, \tag{5.6}$$

$$S_{\hat{\beta},\beta} = \frac{1}{J}\sum_{t=T}^{T+J-1}\left(y_{\hat{\mathbf{F}},t+1|t} - \hat{y}_{\hat{\mathbf{F}},t+1|t}\right)^2. \tag{5.7}$$

Table 1: Small sample bias and standard errors for the estimated factors $\tilde{F}_{t,1}$, for $t = [T/2], T$ and $r = 1$

		PCE				CnPCE			
		$\tilde{F}_{[T/2],1}$		$\tilde{F}_{T,1}$		$\tilde{F}_{[T/2],1}$		$\tilde{F}_{T,1}$	
T	N	bias	std	bias	std	bias	std	bias	std
50	50	-0.018	0.190	-0.092	0.163	-0.024	0.037	-0.139	0.097
	100	0.001	0.179	-0.207	0.092	0.033	0.031	-0.025	0.008
	150	-0.155	0.136	0.293	0.137	-0.211	0.103	0.353	0.166
100	50	-0.004	0.118	0.008	0.092	-0.046	0.025	0.121	0.015
	100	-0.128	0.102	-0.109	0.106	-0.120	0.079	-0.114	0.054
	150	-0.168	0.105	-0.049	0.115	-0.126	0.063	-0.013	0.053
150	50	-0.007	0.089	0.026	0.078	-0.032	0.053	0.070	0.081
	100	0.018	0.097	-0.113	0.086	0.054	0.022	-0.180	0.032
	150	0.093	0.062	0.031	0.065	-0.000	0.020	-0.065	0.021

The results are for the sampling distribution of $\tilde{F}_t = \mathcal{J}^{-1}\hat{F}_t$, $\mathcal{J} = \mathcal{H}$ for CnPC and $\mathcal{J} = H$ for PCE. The thresholding parameter M is chosen by a 10-fold cross-validation.

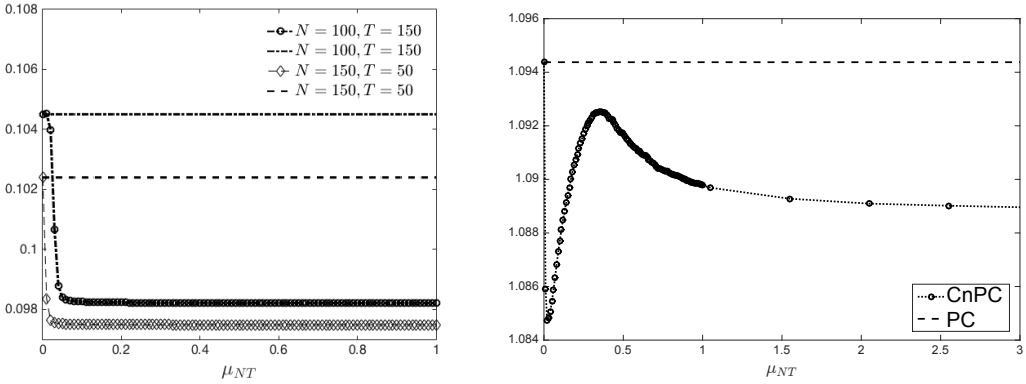
The statistic in (5.6) measures the loss in forecast accuracy due to F_t being unobserved and estimated. If the estimated factors are consistent and span the same space as the true factors, the difference in forecasting performance of the two predictors $\hat{F}_{t,N}$ and F_t^0 should be negligible and $S_{\hat{y}^{\hat{\mathbf{F}}},\hat{y}^{\mathbf{F}^0}}$ close to one. The larger is $S_{\hat{y}^{\hat{\mathbf{F}}},\hat{y}^{\mathbf{F}^0}}$, the closer are the 'diffusion index' forecasts to those generated by the unfeasible forecasts computed based on the true observed factors. The statistic in (5.7) assesses the accuracy of the 'diffusion index' forecasts relative to the conditional mean forecasts which requires only estimation of F_t. Smaller values of $S_{\hat{\beta},\beta}$ are desirable.

5.3 Simulation results

Case of threshold M selected by cross-validation

Table 1 reports the small sample bias and sample standard deviation of the estimated factors $\tilde{F}_{tj}$ for $j = 1$. The table reports results for two arbitrary choices of t: $[T/2]$ and T. The number of factors in this design experiment is $r = 1$. In all Monte Carlo results, the number of factors r is assumed to be known and is not estimated. The threshold M is selected with a 10-fold cross-validation with minimum average MSE, of estimated factors, as a risk function. Overall, the proposed CnPCEs suffer some small

Figure 1: Empirical mean-squared-errors accuracy of the CnPC estimators of common factors

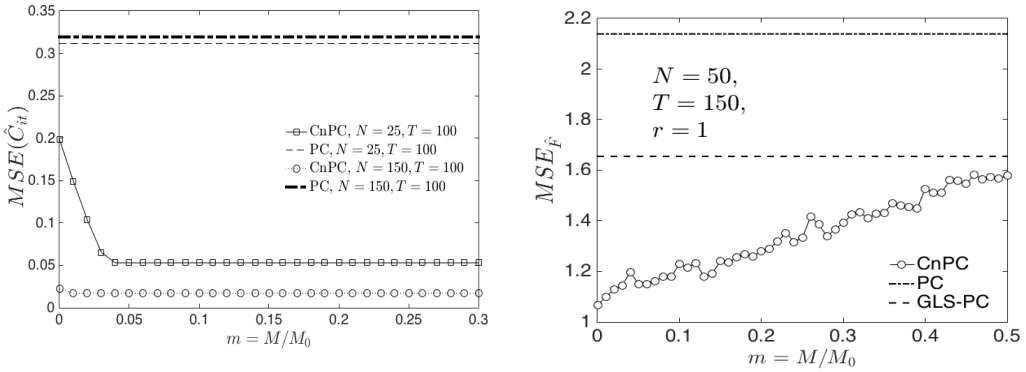


(a) Case $r = 1$ (b) Case $r = 4$

Note: The MSEs are computed for the rotated estimated factor matrix $\tilde{F}_t = \mathcal{J}\hat{F}_t$. The shrinkage factor M is set equal to the value for which the constrained problem has the same solution as its dual penalized PC regression. The graphs show the effect of μ_{NT}.

sample bias, compared to the ordinary PCEs, especially for small N large T panels. These results are expected since the CnPCEs have a slower rate of convergence due to the penalization factor. There is however an appreciable gain in efficiency as measured with smaller sample standard errors results for the CnPCEs, especially for panels with large N. Figure 1 displays the sample (empirical) mean-squared-errors (MSEs) (5.3) for the rotated factors $\widetilde{F}_t$, estimated using the CnPC over a grid of values for the regularization parameter μ_{NT} and for a given M. The results shown are equivalent to the penalized PC estimator that solves (3.8). The results for the PCEs are displayed with a dashed line. The left panel is for the case with one true factor and the right panel is for the case of two factors in the population model. As expected, the proposed technique with $\mu_{NT} = 0$ gives the same factors' accuracy in terms of MSEs as the standard PC method. As the penalization increases, the MSEs for model with $r = 1$ decrease sharply. For data generating process with $r = 4$, the MSEs of $\tilde{F}_t$ also reach a stable value, that is significantly lower than the empirical MSE of the PCEs, after some dynamics for small μ_{NT}. The relationship between MSEs and μ_{NT} is not monotonic. Figure (2) displays the MSEs for the CnPC estimates of the common factors $\hat{\mathbf{F}}$ in the right panel and the common components, $\hat{C} = \hat{\mathbf{\Lambda}}\hat{\mathbf{F}}'$, in the left panel. In the right

Figure 2: Empirical mean squared errors of the CnPC estimators of the common factors $\hat{F}_t$ and the common components $\hat{C}_{it}$, where $i = 1, \cdots, N$ and $t = 1, \cdots, T$, and $r = 1$.



(a) Common Components $MSE_{\hat{\mathbf{C}}}$ (b) Common Factors $MSE_{\hat{\mathbf{F}}}$

panel, we extend the comparative analysis to a GLS-type PCE as in Choi [2012]. Note that a GLS-PCE method requires inverting the sample covariance matrix. Therefore a GLS-PC estimation is restricted to panels with $N < T$. Overall, both estimators bring significant improvement over the ordinary PCEs, as measured with lower MSEs results for the factors. The MSE for the GLS-PCE is represented with a straight dashed-line, as it is not dependent on the thresholding factor m. The GLS-PCE delivers a 22% decrease in the MSE of the estimated factors. The striking result is however the improved performance of the CnPC estimator with consistent gains in MSE over the GLS-PCE, ranging from 5% to 35% depending on m.

Consider the case of the common factors in the left panel. Accuracy of the CnPC, as measured by the MSE, increases as N becomes large. This is not the case for the PC estimators where the gains in forecasting accuracy are very small.

5.3.1 The case of M indexed path

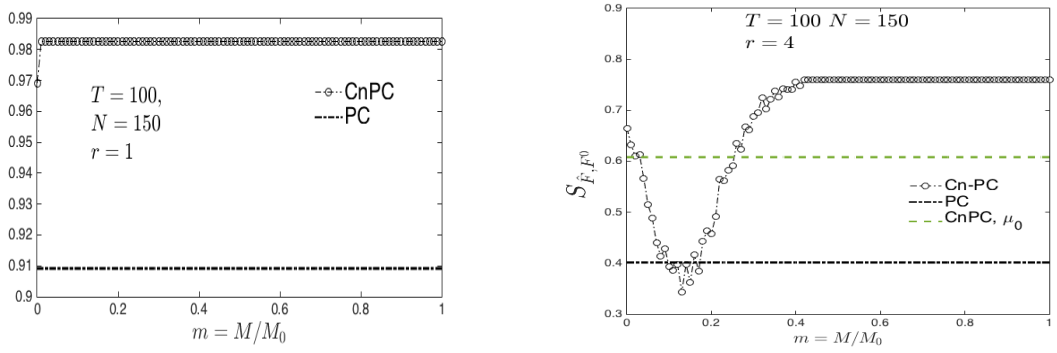
Figure 3 displays the path of the statistic $S_{\hat{F},F^0}$ indexed by $m = M/M_0$. Note that in this experiment, the CnPC estimator $\hat{F}_t$ and $\hat{\mu}_{NT}$ are jointly estimated. The CnPC estimator results are shown in a circle-dot dashed line and the PCEs in a bold dashed line. The left panel plots the results for a DGP with four factors, and the right panel

Table 2: Accuracy of estimated common factors $\hat{\mathbf{F}}$ in spanning the true factor space of $\mathbf{F}^0$: $S_{\hat{\mathbf{F}},\mathbf{F}^0}$ and $MSE_{\hat{\mathbf{F}}}$

		$S_{\hat{\mathbf{F}},\mathbf{F}^0}$			$MSE_{\hat{\mathbf{F}}}$		
T	N	PC%	CnPC%	GLS-PC%	PC	CnPC	GLS-PC
100	25	13.0	31.9	43.4	2.17	2.16	2.13
	50	12.0	38.2	15.7	1.84	1.76	1.77
150	50	10.3	33.7	18.5	1.78	1.95	1.97
	100	10.5	58.0	7.8	1.83	1.89	1.94
55	50	24.2	50.5	7.2	1.94	1.95	1.75
50	25	26.7	34.1	25.2	1.96	1.36	2.02

plots the results for a one factor model. The right panel shows that the PCE is doing a good job in mapping the space of the true factors with accuracy results in the 90% range. However, the CnPC estimator has a clear advantage with values in the 97% range. For the four-factor case on the left, the estimated factors span less perfectly

Figure 3: Accuracy of CnPC estimators of the common factors $\hat{F}$: $S_{\hat{F},F^0}$



(a) Case $r = 1$ (b) Case $r = 4$

the true space of the true factors. The explained variation in the true factors for the PCEs is considerably lower (in the 40% range). On the other hand, the results for the CnPC estimator are promissing with accuracy values ranging up to %75, thus improving the ability of the estimated factors estimates to span the true factor space. The plot also suggests that the relation between M and $S_{\hat{F},F}$ is not monotonic with

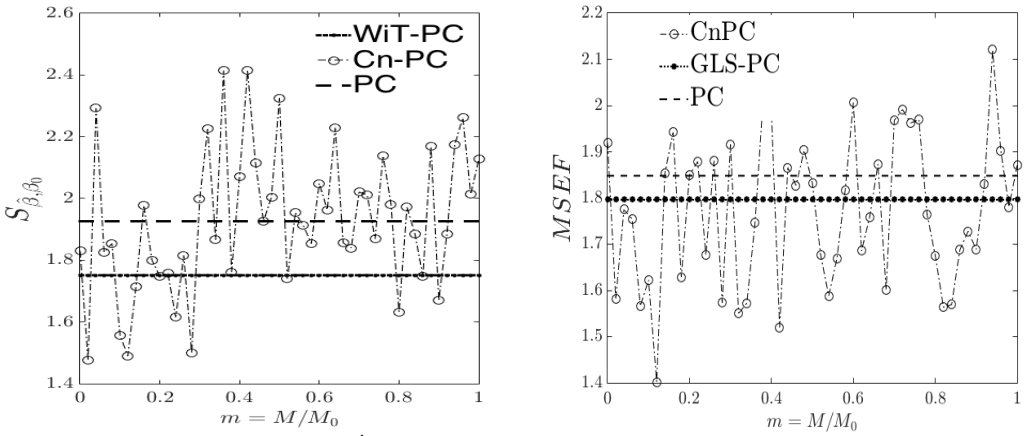
some dynamics in the low range of the grid. These plots clearly show that the CnPCE sample performance is affected by the thresholding criteria. For example, the dashed-line in the plot represents the results for a fixed value of M corresponding to $\mu = \mu_0$, the initialization value specified in the algorithm. These plots can be used as graphical tools for eliciting M. The *ridge-trace* plot is a similar strategy that is used in the context of ridge regression [Hoerl et al., 1975]. Such plots provide a visual assessment of the effect on the regression coefficient of the choice of the ridge regularization parameter, thus allowing the analyst to make a more informed decision. Similarly, the selected M using such strategy would correspond to the value at which the statistic of interest stabilizes.

In Table 2, the estimator GLS-PC refers to Choi [2012] estimator that uses a PCE's sample covariance matrix to compute a feasible generalized PC efficient estimator. The PCE is very inaccurate in terms of $S_{\hat{F},F^0}$. The GLS-PC performs better in case of T large and N small. However, as N becomes larger, GLS-PCE becomes less accurate. When N is large, GLS-PC performs poorly with S_{F,F^0} considerably lower than the ones for the PC and CnPC estimators. The low accuracy of GLS-PCE can be explained by the poor accuracy and unstable sample covariance matrix when N is large and close to T.

Sample correlations

Similar to the ridge-trace plot which graphically shows the effect of the shrinkage parameter on the coefficients of a linear regression model, one can look at the effect of the threshold M on the sample cross-section correlations $|\hat{\tau}_{ij}|$. We use this strategy to select the threshold M for the results in this section. Figure 6 shows histograms of the sampling distribution of $\hat{\Omega}_{ij}$ for an arbitrary selection of values for i and j. We select cases where $\Omega^0_{ij} = 0$ and $\Omega^0_{ij} \neq 0$ in the population model. The dotted vertical line marks the true population value. The CnPC estimators are shown in the black-color

Figure 4: Accuracy of the Diffusion Index forecasts



(a) Accuracy of $\hat{\beta}$ (b) Accuracy of $\hat{y}_{t+12|\hat{F}_t}$

histogram.

In the top two panels, the results show that for the PCE's estimates of the sample correlations, the distribution is almost symmetric around zero and fat-tailed. The CnPCE's estimates of these correlations are much smaller and concentrated around a small average value. This observation is independent of the true population value. The important finding here is that the CnPC estimator shrinks the average absolute value of these correlations.

These finding are indeed surprising given that empirically we observe a shrinkage at the level of each correlation, while the estimation sets up the constraints at the average level. The CnPC estimator's correlations are shrunk relative to the PCEs. This reduction in the size of the correlations is less significant for the case of $N = T = 150$, although the spread is still smaller.

Figure 5 shows the sampling distribution of maximum average cross-correlation $\hat{\tau}^*$. The results show that overall, the estimated $\hat{\tau}^*$ based on the CnPC estimator are lower than those based on PCEs. In the first panel with $N = 50$ and $T = 100$, $\hat{\tau}^*$ for the CnPC estimator support ranges from 0.02 to 0.46, while for the PCEs the range starts at 0.47 and goes to 0.66.

Figure 5: Sampling distribution of $\hat{\tau}^*$

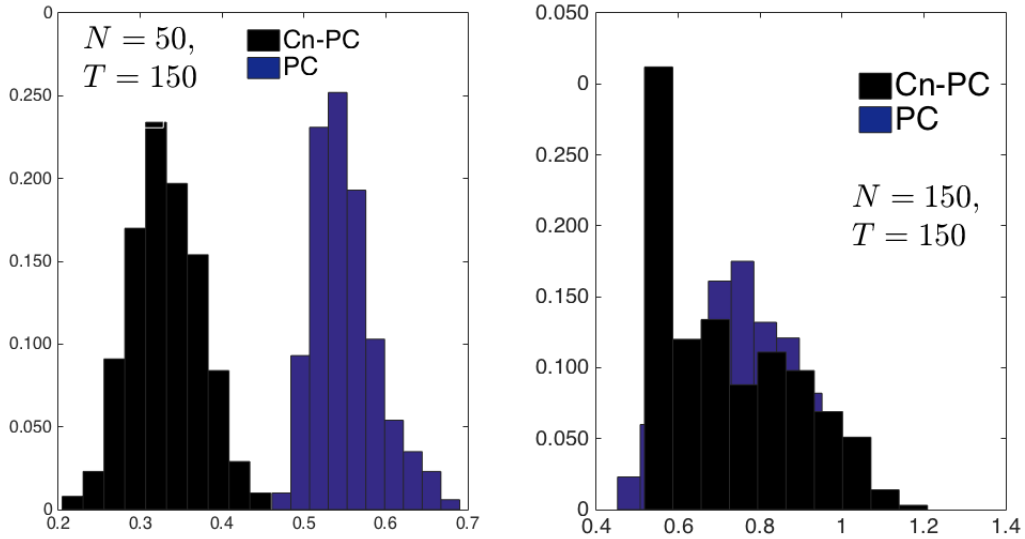
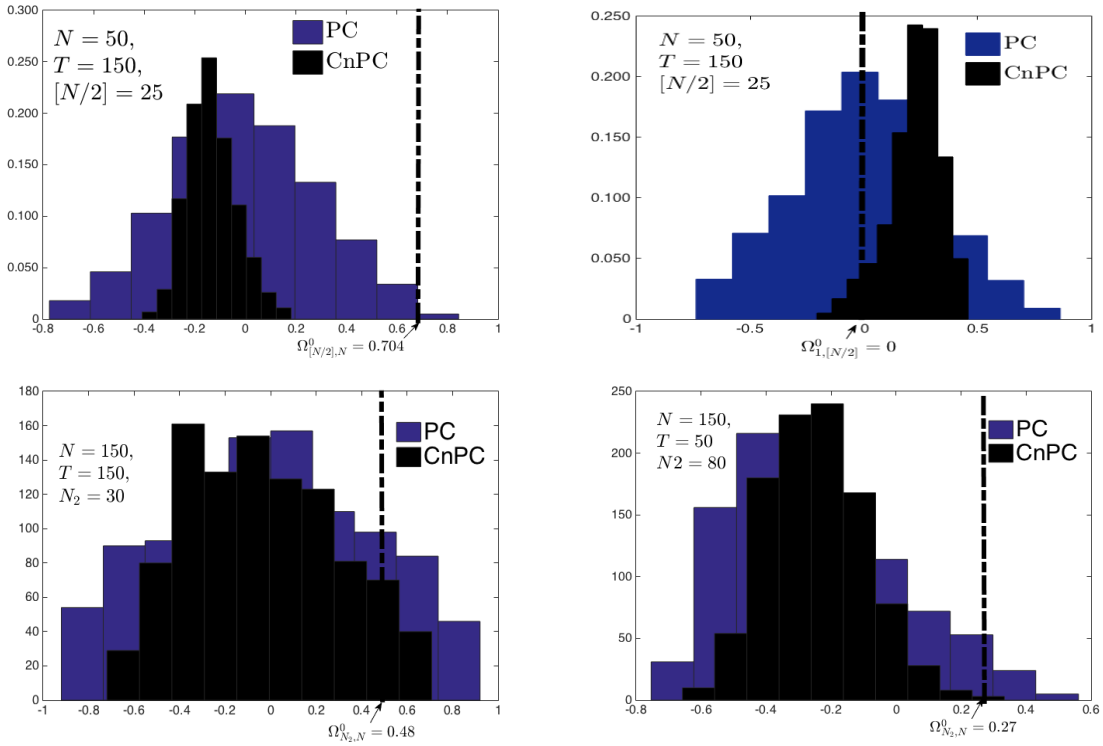


Figure 6: Distribution of $\hat{\Omega}_{i,j}$, the elements of the sample covariance matrix $\hat{\Omega}$.



These results depend on the panel dimension. For $T = N = 150$, the results are less promising, although the distribution of $\hat{\tau}^*$ is skewed to the left, favoring lower values.

Simulated forecasts

Figure 4 displays the statistics $S_{\hat{y},y}$ and $S_{\hat{\beta},\beta_0}$ indexed by $m = M/M_0$. The dotted-dashed circle line plots the results for the CnPC estimator, while the benchmark PCEs are shown in the straight-dashed line. The plot also shows the results for the weighted-PC estimator [Boivin and Ng, 2006], which uses as weights w_{iT} equal to the inverse of $N^{-1}\sum_{j=1}^{N}|\hat{\Omega}_{ij}|$ for each error e_{it} in the PC objective function. The results correspond to a panel with $T = 120$ and $N = 130$, to reflect the panel dimensions that are encountered in macroeconomic forecasting and arbitrage pricing applications. The plots correspond to averages over 1000 replications.

As expected, the weighted-PC estimator outperforms the PCEs with smaller values of $S_{\hat{y},y}$ and $S_{\hat{\beta},\beta_0}$. The shrinkage factor $M = m \cdot M_0$ matters for the performance of the CnPC. Unlike the results we have documented earlier with respect to the accuracy of the factors, there is no pattern to the relationship between M and the accuracy of the diffusion index forecasts. But the results show that, for small values of m, the CnPC can outperform the weighted-PC by sizable margins.

6 Application

This section applies the CnPC estimator to a forecasting experiment for the U.S. Index of Industrial Production (IPS10) and Consumer Price Index (PUNEW) using the dataset provided by Stock and Watson [2002a]. The data include real variables such as sectoral industrial production, employment and hours worked; and nominal variables such as consumer and price indexes, wages, money aggregates, stock prices and exchange rates. The data series are transformed to achieve stationarity: monthly

Table 3: Pseudo-out-of-sample mean squared forecasts errors for US inflation and industrial production

		IPS10				PUNEW			
		$r = 10$		$r = 5$		$r = 10$		$r = 5$	
h=12		PC	CnPC	PC	CnPC	PC	CnPC	PC	CnPC
1970-2002	$MSFE$	0.51	0.51	0.52	0.50	0.64	0.62	0.57	0.57
	Var	0.85	0.85	0.66	0.66	0.53	0.53	0.60	0.60
1970-1985	$MSFE$	0.32	0.31	0.31	0.31	0.43	0.40	0.38	0.38
	Var	0.95	0.94	0.75	0.75	0.45	0.45	0.56	0.56
1985-2002	$MSFE$	1.09	1.08	1.13	1.11	1.65	1.63	1.46	1.40
	Var	0.53	0.50	0.39	0.43	0.87	0.85	0.77	0.75

		IPS10				PUNEW			
		$h = 1$		$h = 4$		$h = 1$		$h = 4$	
r=7		PC	CnPC	PC	CnPC	PC	CnPC	PC	CnPC
1970-2002	$MSFE$	0.72	0.70	0.57	0.57	0.78	0.75	0.67	0.67
	Var	0.42	0.38	0.56	0.56	0.27	0.27	0.37	0.37
1970-1985	$MSFE$	0.66	0.61	0.49	0.49	0.75	0.71	0.56	0.55
	Var	0.46	0.43	0.56	0.56	0.26	0.25	0.42	0.41
1985-2002	$MSFE$	0.86	0.86	0.86	0.86	0.82	0.82	0.97	0.97
	Var	0.28	0.28	0.54	0.54	0.28	0.28	0.25	0.25

growth rates for real variables (e.g. industrial production, sales) and first differences for variables already expressed in rates (e.g. unemployment rate, capacity utilization).

The dataset comprises monthly observations from 1959:01 to 2003:12 and 131 time series. The sample is divided into an in-sample portion of size $T = 120$ (1959:01 to 1969:12) and an out-of-sample evaluation portion with the first date as December 1970 and the last date as December 2003. There are a total of $J = 397$ out-of-sample evaluation points split into pre- and post-1985 periods, with a cut-off date of December 1984. This cut-off date has economic significance due to a change in monetary policy to inflation targeting, and thus the well-known moderation period. The models and parameters are re-estimated and the 12-step-ahead forecasts are computed for each month $t = T + 12, \cdots T + 12 + J - 1$ using a rolling window scheme that uses the most recent 10 years of monthly data, that is data indexed $t - 12 - T + 1, \cdots, t - 12$.

In this empirical example, the CnPC estimator is computed using a threshold parameter M that is chosen using a 10-fold cross-validation.

Table 3 reports the mean squared forecasts error (MSFE) relative to the random walk and the variance (var) of the forecasts, relative to the variance of the series to be forecast. We consider three sample periods and consider different values for the forecast

horizon h. The number of factors r is selected using Bai and Ng's (2002) information criterion IC_{p_1}, which returns an estimate of $\hat{r} = 7$. We also show results for arbitrary values of $r = 5, 10$.

We observe that overall, there are gains in the out-of-sample forecasts accuracy as measured by average mean-squared-forecast-errors. These gains depend on the sample period and on the target series. Generally, the gains can be as high as a 6% decrease in the pseudo-out-of-sample mean-squared forecast errors.

Consumer Price Index forecasts appear to benefit the most from incorporating dependence features using the CnPC estimators of the predictors $\hat{F}_t$. These benefits are more appreciable during the post-moderation period 1985–2002. This result is supported by previous findings in the literature. During this period, predictability of the price and output series is problematic partly because of the instabilities in the data, and partly because of the FED's monetary policy of inflation targeting.

7 Conclusion

This paper proposes a novel PC-based method for incorporating the features of cross-correlation in the data in large factor models. The method allows for approximate factor structure in the sense of Chamberlain and Rothschild [1983], and embeds the assumption of bounded cross-sectional dependence to solve a constrained PC problem. This constrained estimation is easily implemented within the existing classical PC analysis. The method does not require inverting a large covariance matrix and works through a shrinkage mechanism applied to the sample average of cross-sectional correlations.

We provide convergence results of the estimated factors to a space that spans the true factors. The convergence rate is slower than in the asymptotic classical PC method and depends on a regularization parameter. The method is computationally equivalent

to a PC estimation applied to a regularized form of the data matrix.

The simulation analysis shows that the CnPC estimator is generally more accurate than the PCE and GLS-PCE for large N panels. Applied to real data, the results suggest that sizable improvements in the accuracy of the estimated factors, in spanning the true factors space, do not always lead to quantitaively similar improvements in the forecasts' accuracy of the diffusion indexes, and that the results greatly depend on the target series and on the forecast's horizon.

References

Ahn, S. C., Horenstein, A. R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81 (3), 1203–1227.

Alessi, L., Barigozzi, M., Capasso, M., 2010. Improved penalization for determining the number of factors in approximate factor models. Statistics and Probability Letters 80 (23), 1806 – 1813.

Amengual, D., Watson, M. W., 2007. Consistent estimation of the number of dynamic factors in a large n and t panel. Journal of Business & Economic Statistics 25 (1), 91–96.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–172.

Bai, J., Li, K., 2012. Statistical analysis of factor models of high dimension. Annals of Statistics 40, 436–465.

Bai, J., Liao, Y., 2016. Efficient estimation of approximate factor models via penalized maximum likelihood. Journal of Econometrics.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bai, J., Ng, S., 2003. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. Mimeo, University of Michigan.

Bai, J., Ng, S., 2007. Determining the number of primitive shocks in factor models. Journal of Business and Economic Statistics, 52–60.

Bai, J., Ng, S., 2013. Principal components estimation and identification of static factors. Journal of Econometrics 176 (1), 18–29.

Bai, J., Ng, S., 2017. Principal components and regularized estimation of factor models. arXiv:1708.08137v2.
URL `https://arxiv.org/abs/1708.08137`

Bickel, P. J., Levina, E., 2008a. Covariance regularization by thresholding. Annals of Statistics 36, 2577—-2604.

Bickel, P. J., Levina, E., 2008b. Regularized estimation of large covariance matrices. Annals of Statistics 36, 199—-227.

Boivin, J., Ng, S., 2006. Are more data always better for factor analysis? Journal of Econometrics 132, 169–194.

Breitung, J., Tenhofen, J., 2011. GLS estimation of dynamic factor models. Journal of the American Statistical Association 106 (495), 1150–1166.

Brown, S. J., 1989. The number of factors in security returns. Journal of Finance 44, 1247–1262.

Cai, T., Liu, W., 2011. Adaptive thresholding for sparse covariance matrix estimation. Journal of the American Statistical Association 106, 672—684.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and mean-variance analysis in large asset markets. Econometrica 51, 1305–1324.

Choi, I., 2012. Efficient estimation of factor models. Econometric Theory 28, 274–308.

Connor, G., Korajczyk, R. A., 1989. An intertemporal equilibrium beta pricing model. Review of Financial Studies 2 (3), 255–289.

Connor, G., Korajczyk, R. A., 1993. A test for the number of factors in an approximate factor model. Journal of Finance XLVIII (4), 1263–1291.

De Mol, C., Giannone, D., Reichlin, L., 2008. Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? Journal of Econometrics 146, 318–328.

Doz, C., Giannone, D., Reichlin, L., 2012. A quasi maximum likelihood approach for large approximate dynamic factor models. Review of Economics and Statistics (REStat) 94 (4), 1014–1024.

Dürre, A., Vogel, D., Fried, R., 2015. Spatial sign correlation. Journal of Multivariate Analysis 135, 85–105.

Fan, J., Liao, Y., Liu, H., 2016. An overview of the estimation of large covariance and precision matrices. The Econometrics Journal 19 (1), C1–C32.

Fan, J., Liao, Y., Mincheva, M., 2011. High-dimensional covariance matrix estimation in approximate factor models. The Annals of Statistics 39 (6), 3320–3356.

Fan, J., Liao, Y., Mincheva, M., 2013a. Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75 (4).

Fan, J., Liao, Y., Mincheva, M., 2013b. Large covariance estimation by thresholding principal orthogonal complements with discussion. Journal of the Royal Statistical Society, Series B Statical Methodology 75, 603–680.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2004. The generalized dynamic factor model: consistency and rates. Journal of Econometrics 119, 231–245.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. The generalized dynamic factor model: One sided estimation and forecasting. Journal of the American Statistical Association 100, 830–840.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2009. Opening the black box: Structural factor models with large cross-sections. Econometric Theory 25 (05), 1319–1347.

Forni, M., Hallin, M., Reichlin, L., 2000. The generalized dynamic factor model: Identification and estimation. Review of Economics and Statistics, 540–554.

Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association 102 (478), 603–617.

Hoerl, A. E., Kennard, R. W., Baldwin, K. F., 1975. Ridge regression: Some simulations. Communications in Statistics 4 (2), 105—123.

Kapetanios, G., 2010. A testing procedure for determining the number of factors in approximate factor models with large datasets. Journal of Business and Economics Statistics 28, 397–409.

Lam, C., Fan, J., 2009a. The Annals of Statistics 37 (6B), 4254–4278.

Lam, C., Fan, J., 2009b. Sparsistency and rates of convergence in large covariance matrix estimation. The Annals of Statistics 37 (6B), 4254–4278.

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal Multivariate Analysis 88, 365–411.

Ledoit, O., Wolf, M., 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. The Annals of Statistics 40 (2), 1024—-1060.

Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. Econometrica 77 (5), 1447–1479.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. The Review of Economics and Statistics 92 (4), 1004–1016.

Osborne, M. R., Presnell, B., Turlach, B. A., 2000. On the LASSO and its dual. Journal of Computational and Graphical Statistics 9 (2), 319–337.

Stock, J., H., Watson, M., W., 1998. Diffusion indexes. NBER, Working Papers 6702.

Stock, J. H., Watson, M. ., 2006. Forecasting with many predictors. In Handbook of Economic Forecasting 1, 551–554.

Stock, J. H., Watson, M. W., 2002a. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Stock, J. H., Watson, M. W., 2002b. Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20 (2), 147–162.

Stock, J. H., W. M. W., 2005. An empirical comparison of methods for forecasting using many predictors. Manuscript. Princeton University.